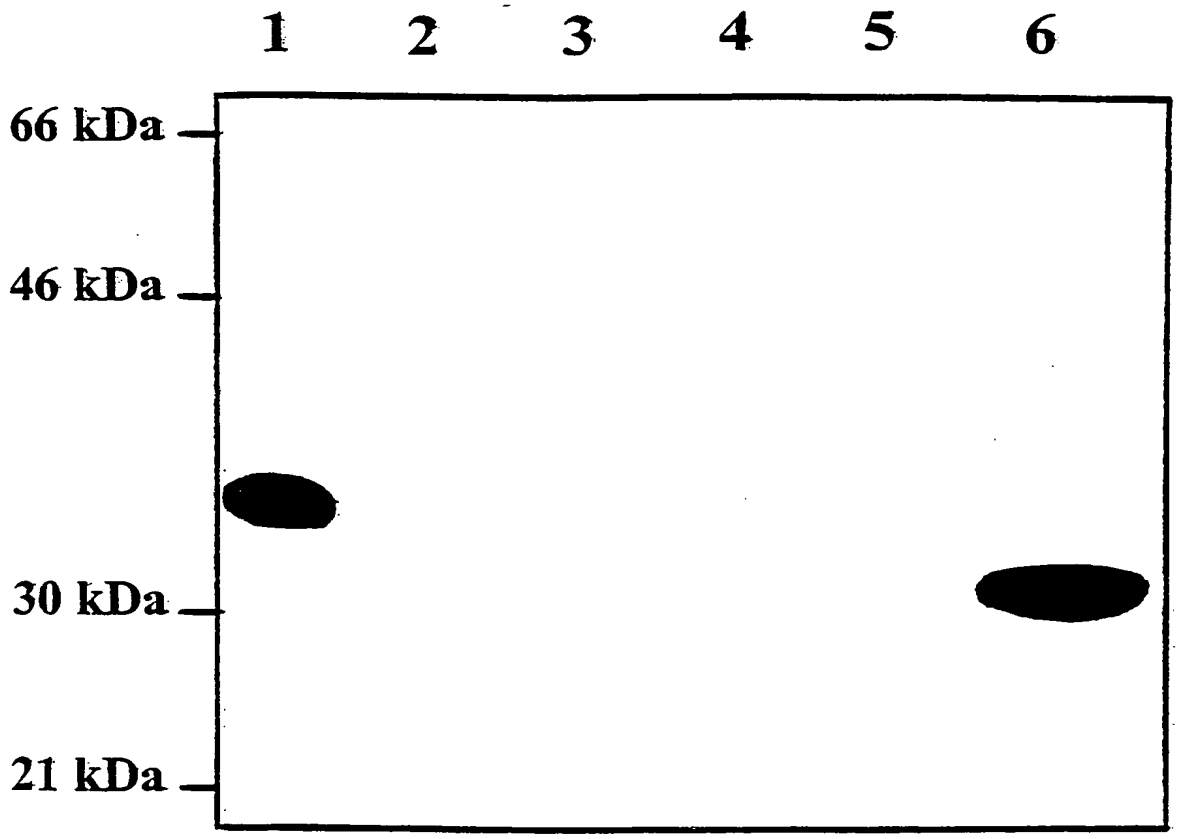
Fig. 1



Lanes:
1. 2.5 ng HT-IN
2. pCEP-4
3. pCEP-IN
4. pCEP-IN-CTE
5. pCEP-IN-RRE + pEF-cRev
6. pCMV-IN^S

Fig. 2A

```
ATCACTAGCA ACCTCAAACA GACACCATGG GATTCCTGGA CGGCATTGAC AAGGCTCAGG AGGAGCACGA GAAGTACCAC TCGAATTGGC GGGCCATGGC  100
                                M  G  F  L  D  G  I  D  K  A  Q  E  E  H  E  K  Y  H  S  N  W  R  A  M  A

                                                    NheI
CTCCGACTTC AACCTGCCAC CCGTCGTCGC TAAGGAGATC GTTGCTAGCT GCGACAAGTG CCAGCTGAAA GGCGAGGCTA TGCACGGGCA GGTTGATTGC  200
 S  D  F  N  L  P  P  V  V  A  K  E  I  V  A  S  C  D  K  C  Q  L  K  G  E  A  M  H  G  Q  V  D  C

TCTCCCGGCA TCTGGCAGCT CGACTGTACT CACCTGGAGG GCAAGGTCAT CCTGGTCGCC GTGCACGTGG CCTCTGGTTA CATCGAGGCT GAGGTCATCC  300
 S  P  G  I  W  Q  L  D  C  T  H  L  E  G  K  V  I  L  V  A  V  H  V  A  S  G  Y  I  E  A  E  V  I  P

      PstI
CTGCAGAGAC TGGCCAGGAG ACTGCCTATT TCCTGCTGAA ACTGGCCGGC CGGTGGCCTG TGAAGACAGT GCACACAGAT AACGGCTCCA ACTTCACCTC  400
 A  E  T  G  Q  E  T  A  Y  F  L  L  K  L  A  G  R  W  P  V  K  T  V  H  T  D  N  G  S  N  F  T  S

                                                BamHI
CACCACTGTG AAGGCTGCCT GCTGGTGGGC TGGGATCAAG CAGGAGTTCG GGATCCCCTA TAACCCACAG TCTCAGGGCG TGATCGAATC CATGAACAAG  500
 T  T  V  K  A  A  C  W  W  A  G  I  K  Q  E  F  G  I  P  Y  N  P  Q  S  Q  G  V  I  E  S  M  N  K

GAGCTGAAGA AGATCATCGG CCAGGTTCGG GACCAGGCAG AGCACCTGAA GACTGCAGTG CAGATGGCCG TGTTCATCCA CAACTTCAAG CGAAAGGGCG  600
 E  L  K  K  I  I  G  Q  V  R  D  Q  A  E  H  L  K  T  A  V  Q  M  A  V  F  I  H  N  F  K  R  K  G  G

                  NaeI
GCATCGGTGG CTACTCAGCC GGCGAGCGGA TCGTGGACAT CATCGCCACT GACATCCAGA CCAAAGAGCT GCAGAAGCAG ATCACCAAGA TCCAGAACTT  700
 I  G  G  Y  S  A  G  E  R  I  V  D  I  I  A  T  D  I  Q  T  K  E  L  Q  K  Q  I  T  K  I  Q  N  F

                                                NarI
CCGTGTGTAC TACCGGGACT CCCGGGACCC TGTGTGGAAG GGCCCTGCCA AGCTGCTGTG GAAGGGCGAG GGCGCCGTGG TCATTCAGGA CAACTCTGAC  800
 R  V  Y  Y  R  D  S  R  D  P  V  W  K  G  P  A  K  L  L  W  K  G  E  G  A  V  V  I  Q  D  N  S  D

ATCAAGGTTG TGCCCAGGCG CAAGGCCAAG ATTATCCGGG ACTACGGCAA GCAGATGGCT GGCGACGACT GTGTGGCCTC TCGTCAAGAT GAGGACTAAG  900
 I  K  V  V  P  R  R  K  A  K  I  I  R  D  Y  G  K  Q  M  A  G  D  D  C  V  A  S  R  Q  D  E  D

TCCAACTACT AAACTGGGGG ATATTATGAT                                                                            930
```

SUBSTITUTE SHEET (RULE 26)

Fig 2B

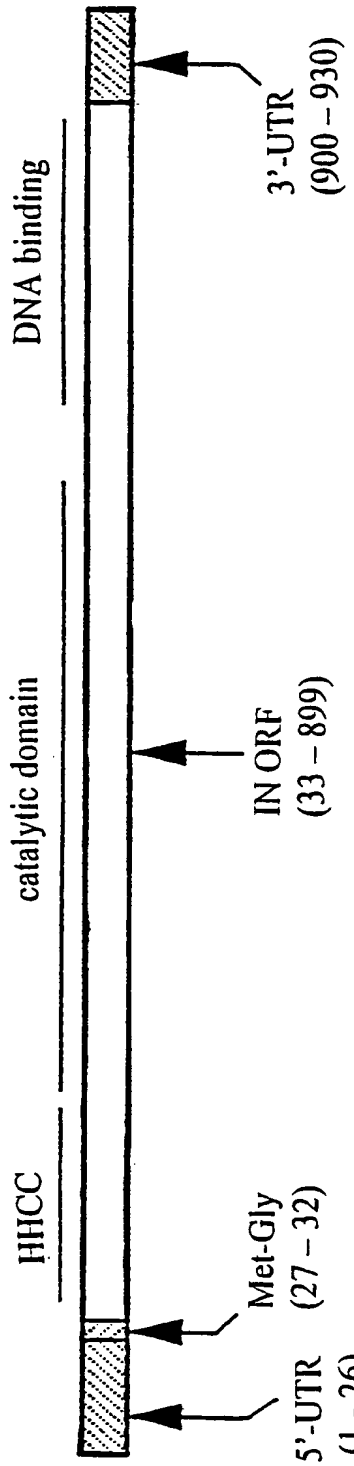
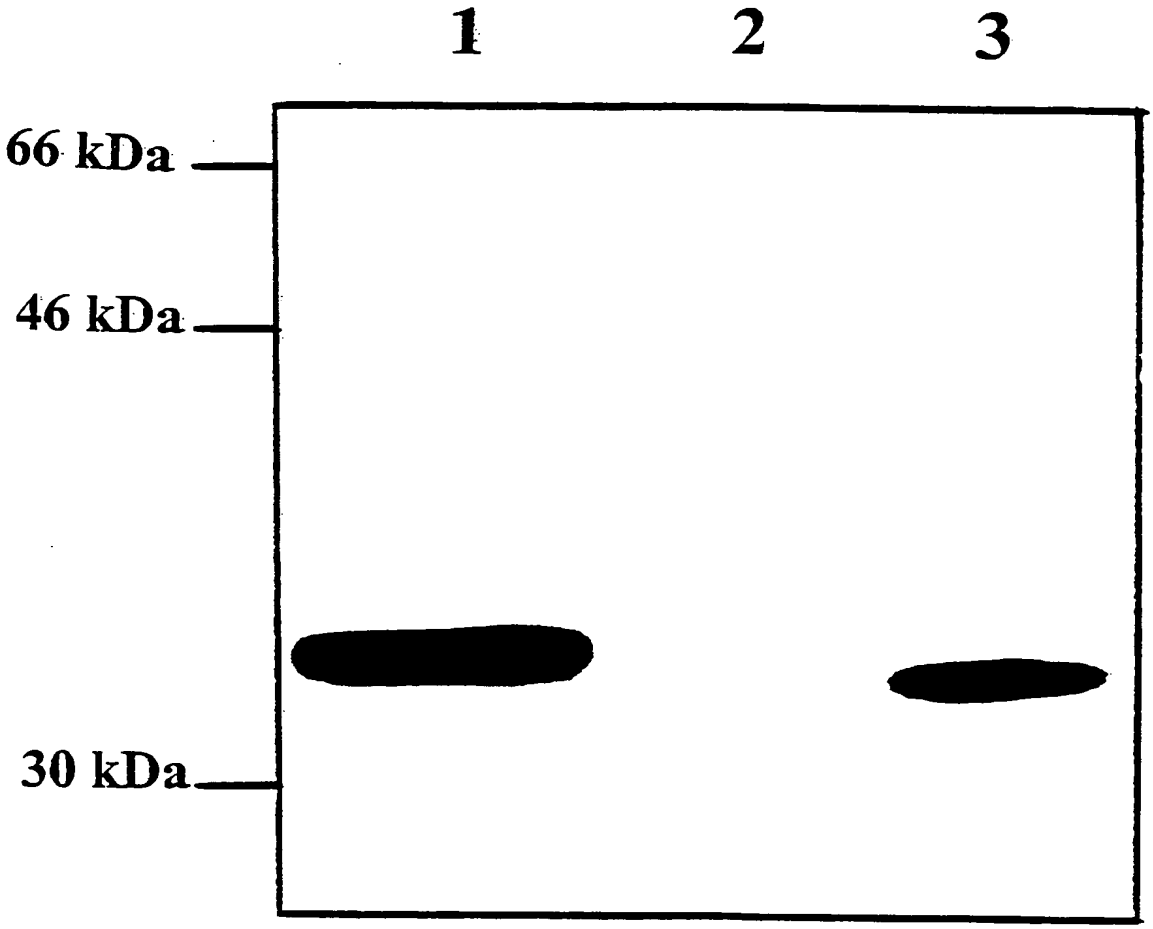

HHCC
catalytic domain
DNA binding
5'-UTR (1 - 26)
Met-Gly (27 – 32)
IN ORF (33 – 899)
3'-UTR (900 – 930)

Fig. 3

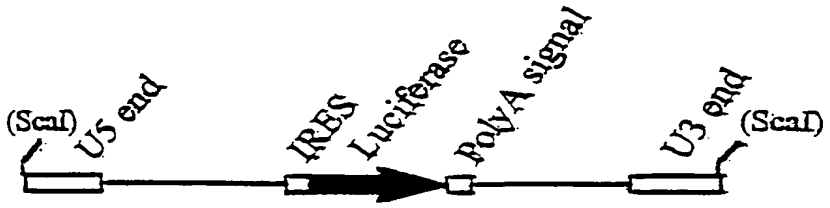


Lanes: 1. 2.5 ng HT-IN
2. 293T
3. 293T-IN^S

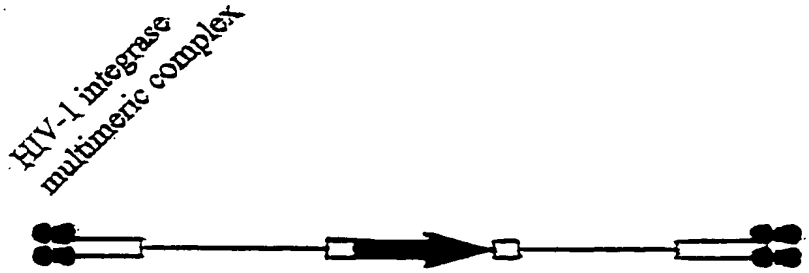
Fig. 4. Principle of DIPR

Detection of integrase activity using a promoterless reporter gene

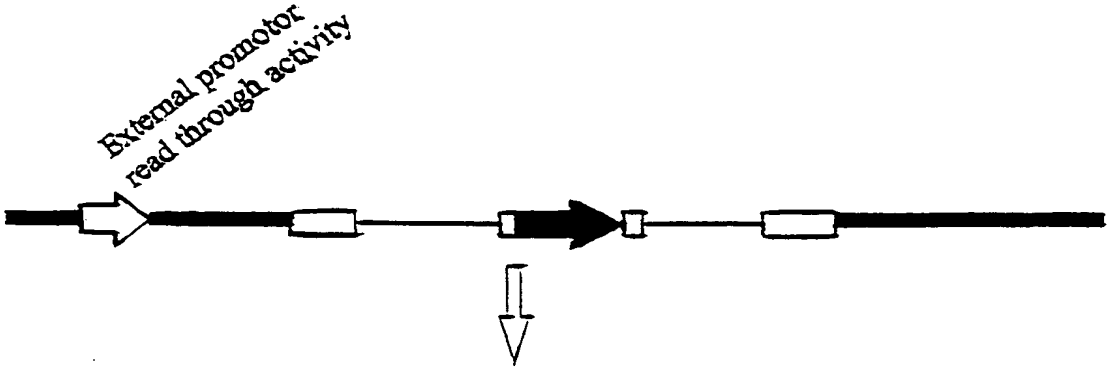
A. Substrate LTR-IRES-Luc (digested with ScaI)



B. Transfection into cells, binding of integrase to U3-U5 ends and cleavage of termini



C. Integration into actively transcribed regions of genomic DNA



Luciferase expression